J Syst Sci Complex (2021) 34: 2048–2068

Distributed Recursive Projection Identification with Binary-Valued Observations*

WANG Ying · ZHAO Yanlong · ZHANG Ji-Feng

DOI: 10.1007/s11424-021-1267-x
Received: 22 July 2021
©The Editorial Office of JSSC & Springer-Verlag GmbH Germany 2021

Abstract This paper investigates a distributed recursive projection identification problem with binary-valued observations built on a sensor network, where each sensor in the sensor network measures partial information of the unknown parameter only, but each sensor is allowed to communicate with its neighbors. A distributed recursive projection algorithm is proposed based on a specific projection operator and a diffusion strategy. The authors establish the upper bound of the accumulated regrets of the adaptive predictor without any requirement of excitation conditions. Moreover, the convergence of the algorithm is given under the bounded cooperative excitation condition, which is more general than the previously imposed independence or persistent excitations on the system regressors and maybe the weakest one under binary observations. A numerical example is supplied to demonstrate the theoretical results and the cooperative effect of the sensors, which shows that the whole network can still fulfill the estimation task through exchanging information between sensors even if any individual sensor cannot.

Keywords Adaptive predictor, binary-valued observations, cooperative excitations, distributed parameter estimation.

1 Introduction

Over the past two decades, the distributed estimation problem over sensor networks has attracted widespread attention and has extensive practical applications, such as target positioning, noise elimination, military surveillance and so on (see [1, 2]). But in most applications, sensors are powered by batteries with finite lifetime, and thus have limited measurement and communication capabilities. Besides, the bandwidth of the communication network is constrained,

WANG Ying · ZHAO Yanlong (Corresponding author) · ZHANG Ji-Feng
Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190*, China; School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing* 100049*, China.*
Email: wangying96@amss.ac.cn; ylzhao@amss.ac.cn; jif@iss.ac.cn.
*The work is supported by National Key R&D Program of China under Grant No. 2018YFA0703800, the National Natural Science Foundation of China under Grant Nos. 61877057 and 62025306, and Open Fund Program of Beijing National Research Center for Information Science and Technology.
◇ *This paper was recommended for publication by Editor XIE Liang-Liang.*

which renders the transmission of vast amounts of real-valued data impractical. Therefore, it is of importance to investigate the distributed estimation problem based on quantized data.

A number of distributed estimation algorithms have been developed (see [3–9]). For instance, the mean-square and steady-state performance analyses of the diffusion least mean square algorithm were studied in [3] with independent and identically distributed (i.i.d.) regressors. And a distributed stochastic approximation algorithm was investigated for the case of uncertain sensing and communication environments, and the convergence properties of the algorithm were discussed under the persistent excitation (PE) condition in [4]. Gan and Liu in [7] presented a distributed stochastic gradient algorithm to estimate the unknown parameter by combining the consensus and diffusion strategies. Besides, Guo and his cooperators in [8] proposed imaginatively a new distributed least-squares (LS) algorithm and obtained some elegant results, such as the convergence of the algorithm under a non-PE condition. Most of the distributed estimation studies, including the above works, are based on accurate communication of sensor networks and accurate measurements of the sensors.

Meanwhile, there are also some studies about distributed estimation problems based on quantized data, which could be divided into two classes. One is the distributed static parameter estimation with quantized communication (see [10–13]), where the sensor has accurate measurements but only can exchange information with its neighbors utilizing quantized communication. For example, the distributed static parameter estimation problem was investigated within the "consensus+innovation" framework utilizing probabilistic quantized communication (see [10, 11]). This problem was considered in [12] over sensor networks under bandwidth constraint and showed that the proposed two-stage averaging-based algorithm achieved the performance of the optimal centralized estimate even if the quantization error variances were not vanishing. Besides, a distributed algorithm, combining a quantized consensus method and the LS approach, was proposed in [13] to study the problem of sensor fusion over networks with asymmetric links, and the performance of the algorithm was analyzed detailedly in terms of unbiasedness and mean square property. The other is the distributed estimation based on quantized observations, where the sensor can exchange information with its neighbors using accurate communication but only has quantized measurements. For example, the distributed parameter estimation problem was studied in [14] for linear systems with quantized observations using normalized least mean square-based consensus algorithm, and the relationship between estimation error and quantized parameter of the uniform quantizer is established.

Actually, binary-valued systems can be found in plentiful important application fields such as engineering areas or bio-medical fields (see [15–17]), so the parameter estimation problem under binary-valued observations has been investigated in many works. Moreover, numerous parameter estimation methods are proposed under binary-valued observations, such as the offline methods including empirical measure method[17, 18], expectation maximization method[19], and the online methods containing recursive projection algorithm[20, 21], recursive LS-type algorithm[22, 23], sign-error type algorithm[24], stochastic approximation type algorithm[25, 26]. It is worth noticing that almost all of the online methods are the stochastic approximation type, i.e., whose gains are scalar forms rather than matrix forms. The key difficulty of studying the matrix-form gain-based algorithms under binary-valued observations is analysing the properties of asymmetric matrix products. It also limits the development of the distributed estimation algorithms under binary-valued observations. Fortunately, Zhang, et al. in [27] overcame the difficulty of matrix analysis by use of a special projection operator and studied the convergence properties of an LS-like algorithm with matrix-form gain under non-PE conditions and binary-valued observations. Enlightened by [27], this paper studies a distributed LS-like algorithm to solve the distributed estimation problem with binary-valued observations.

This paper utilizes matrix inequalities and the martingale estimation theorem to overcome the key difficulty of the distributed estimation problem under binary-valued observations, which is the analysis of the cross items generated by estimation error and binary-valued observations. In contrast to the previous works, the main contributions of this paper can be summarized as follows.

- This paper proposed a distributed recursive projection algorithm to estimate the unknown parameter under binary-valued observations, which is based on a specific projection operator and the diffusion strategies of the neighbor estimates and covariance of regressors. The upper bound of the accumulated regrets of the adaptive predictor is established with no excitation condition. Moreover, the convergence of the proposed algorithm is given under a bounded cooperative excitation condition.
- The cooperative excitation condition in this paper is weaker than persistent excitations used in [20–23], and it might be the weakest one on the system signals under binary observations. Moreover, comparing with the previous studies of parameter estimation under binary measurements (see [17–19, 24–26]), the convergence analysis of the algorithm in this paper does not need the previously imposed periodicity or independence on the system signals. As a benefit, the results in this paper can be applied to feedback control of stochastic systems with only binary-valued observations.
- This paper is inspired by [8] and [27], meanwhile there are some fundamental difficulties. Before guaranteeing the recursions of the Lyapunov function like the method in [27] using the special projection operator, we need to establish the relation between the Lyapunov functions of estimation error after fusion and the one before fusion by some inequalities on the convex combination of positive definite matrices. Besides, the observations of outputs are not accurate but binary-valued compared with [8], hence this paper has to modify the algorithm to deal with binary-valued observations and requires constructing a new martingale difference to estimate the upper bound of the Lyapunov function.

The remainder of this paper is organized as follows. Section 2 introduces some preliminaries on notations, graphs, and observation models. Besides, the distributed recursive projection algorithm is presented. The main results are stated in Section 3, including the convergence of the proposed algorithm and the property of the adaptive predictor. And the proofs of the main results are provided in Section 4. Section 5 uses a numerical example to demonstrate the main results. Section 6 gives concluding remarks and related future works.

2 Problem Formulation

2.1 Some Preliminaries

In this paper, we use $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$ to denote an n-dimensional vector and an $m \times n$-dimensional real matrix, respectively. We use $\|x\| = \|x\|_2$ and $\|A\| = \left(\lambda_{\max}(AA^{\mathrm{T}})\right)^{\frac{1}{2}}$ to denote the Euclidean norms of vector and matrix, respectively, where the notation T is the transpose operator and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the corresponding matrix. Correspondingly, we use $\lambda_{\min}(\cdot)$ to denote the smallest eigenvalue of the matrix, and $\mathrm{tr}(B) = \sum_{i=1}^{n} b_{ii}$ to denote the trace of matrix $B = \{b_{ij}\} \in \mathbb{R}^{n\times n}$. Obviously, if A is a positive semi-definite matrix, then $\mathrm{tr}(A) \geq \|A\|$. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ be two symmetric matrices, then $A \geq B$ means that $A - B$ is a positive semi-definite matrix. The Kronecker product of two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ is defined as $A \otimes B \in \mathbb{R}^{mp\times nq}$. And $|B|$ denotes the determinant of the matrix $B \in \mathbb{R}^{n\times n}$. Besides, the function $I_{\{\cdot\}}$ denotes the indicator function, whose value is 1 if its argument (a formula) is true, and 0, otherwise.

In order to describe the relationship between sensors, an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ is introduced here, where $\mathcal{V} = \{1, 2, \cdots, n\}$ is the set of sensors and $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ is the edge set describing the communication between sensors. The weighted adjacency matrix $\mathcal{A} = \{a_{ij}\} \in \mathbb{R}^{n\times n}$ describes the structure of the graph $\mathcal{G}$, where $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$, and $a_{ij} = 0$, otherwise. We assume the elements of the weighted adjacency matrix $\mathcal{A}$ satisfies $a_{ij} = a_{ji}$, $\forall i, j = 1, 2, \cdots, n$, and $\sum_{j=1}^{n} a_{ij} = 1, \forall i = 1, 2, \cdots, n$. Thus, the adjacency matrix $\mathcal{A}$ is doubly stochastic. Besides, the set of the neighbors of sensor i is denoted as $\mathcal{N}_i = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$, and each sensor can only exchange information with its neighbors in this paper. Moreover, a path of length ℓ of graph $\mathcal{G}$ is a sequence of nodes $\{i_1, i_2, \cdots, i_\ell\}$ satisfying $(i_j, i_{j+1}) \in \mathcal{E}$ for all $1 \leq j \leq \ell - 1$. Graph $\mathcal{G}$ is called connected if for any two agents i and j, there is a path connecting them. The diameter $D(\mathcal{G})$ of graph $\mathcal{G}$ is defined as the maximum value of the distances between any two nodes in graph $\mathcal{G}$.

2.2 Observation Model

Consider a multi-agent network consisting of n sensors, the ith $(i = 1, 2, \cdots, n)$ sensor has the following form

$$\begin{cases} y_{k+1,i} = \varphi_{k,i}^{\mathrm{T}}\theta + d_{k+1,i}, \\ s_{k+1,i} = I_{\{y_{k+1,i} \leq C\}}, \end{cases} \quad k \geq 0, \tag{1}$$

where $\varphi_{k,i} \in \mathbb{R}^m$ is an m-dimensional regressor of the sensor i at time k, $d_{k+1,i} \in \mathbb{R}$ is the noise and $\theta \in \mathbb{R}^m$ is the unknown parameter vector to be identified. And $y_{k+1,i}$ is a scalar output of model i, which only can be measured by binary sensor i, where $s_{k+1,i}$ is the binary-valued observation and C is a fixed threshold of binary sensor i, respectively.

Remark 1 As known, the thresholds of binary sensor have two classes, fixed thresholds and more complicatedly time-varying thresholds. This paper investigates system identification over binary sensor network with fixed thresholds. It is worth mentioning that the similar results can still be obtained with time-varying thresholds, only if there exists a unified bound for these time-varying thresholds. More specifically, we take $C_{k,i}$ as the threshold of sensor i at time k only if there exists a positive constant C_0 such that $|C_{k,i}| \leq C_0$ for all $i = 1, 2, \cdots, n$ and $k \geq 1$, then we can obtain the similar conclusion as this paper.

In order to proceed with the analysis, we introduce some assumptions concerning the priori information of the unknown parameter, the inputs, the noises and the graph.

Assumption 1 There is a known bounded convex compact set $\Omega \subset \mathbb{R}^n$ such that the unknown parameter $\theta \in \Omega$. And denote $\overline{\theta} = \sup_{\eta \in \Omega} \|\eta\|$.

Assumption 2 The regressor $\{\varphi_{k,i}, i = 1, 2, \cdots, n\}$ is $\mathcal{F}_k$-measurable and satisfies

$$\sup_{\substack{k \geq 1 \\ i=1,2,\cdots,n}} \|\varphi_{k,i}\| = \overline{\varphi} < \infty, \text{ a.s.,}$$

where $\{\mathcal{F}_k\}$ is a sequence of nondecreasing σ-algebra and $\overline{\varphi}$ may be a random variable.

Assumption 3 For any $i = 1, 2, \cdots, n$, the noise $\{d_{k,i}\}$ is a sequence of i.i.d. random variables and $d_{k,i}$ is $\mathcal{F}_k$-measurable. Besides, the probability distribution function of $d_{k,i}$ is denoted by $F_i(\cdot)$, and the density function $f_i(x) = \frac{dF_i(x)}{dx}$ satisfies

$$\min_{\substack{i=1,2,\cdots,n \\ x \in [C - \overline{\theta}\overline{\varphi}, C + \overline{\theta}\overline{\varphi}]}} f_i(x) \geq \underline{f} > 0. \tag{2}$$

Remark 2 Though Assumption 2 requires that the regressors are bounded almost surely, it is always possible in the actual control systems, such as saturated control of aircraft in [28]. Besides, Assumption 3 looks complex, but it actually includes the assumption of noises in many other studies on system identification with binary measurements, such as white noise in [19–22]. It is worth noticing that the lower bound $\underline{f}$ in (2) will be used to construct the algorithm in this paper, which is helpful to guarantee the boundness of the Lyapunov function.

Assumption 4 The graph $\mathcal{G}$ is connected.

Remark 3 The assumption on the network topology is natural, avoiding isolated nodes in the network. Besides, from Lemma 8.1.2 in [29], one can find that each entry of the matrix $\mathcal{A}^d$ is positive for $d \geq D(\mathcal{G})$, where $D(\mathcal{G})$ is the diameter of the graph $\mathcal{G}$.

The goal of this paper is to develop a distributed algorithm to estimate the unknown parameter θ, based on the system regressors $\varphi_{k,i}$, the binary-valued observations $s_{k,i}$ and the properties of the noises $d_{k,i}$.

2.3 Distributed Recursive Projection Algorithm

Before presenting the distributed recursive projection algorithm, we give a specific projection operator on the Ω in the following form.

Definition 1 (see [27]) For the convex compact set Ω given in Assumption 1 and any positive definite matrix $Q \in \mathbb{R}^{m \times m}$, the projection operator $\Pi_Q\{\cdot\}$ is defined as

$$\Pi_Q\{\eta\} = \arg\min_{\omega \in \Omega} \|\eta - \omega\|_Q, \quad \forall \eta \in \mathbb{R}^m,$$

where $\|\cdot\|_Q$ is defined as

$$\|\eta\|_Q = \sqrt{\eta^{\mathrm{T}} Q \eta}, \quad \forall \eta \in \mathbb{R}^m.$$

Remark 4 From Lemma 2.1 in [30], the projection operator given by Definition 1 satisfies

$$\|\omega - \Pi_Q\{\eta\}\|_Q \leq \|\omega - \eta\|_Q, \quad \forall \omega \in \Omega,\ \eta \in \mathbb{R}^m.$$

Then, by virtue of this projection operator and the diffusion strategies of the neighbor estimates and covariances of regressors, the distributed recursive projection algorithm is proposed for the system (1) as follows.

Algorithm 1 Distributed Recursive Projection Algorithm

For any given sensor $i \in \{1, 2, \cdots, n\}$, begin with an initial estimate $\widehat{\theta}_{0,i} \in \mathbb{R}^m$ and an initial positive definite matrix $P_{0,i} \in \mathbb{R}^{m\times m}$. The algorithm is recursively defined as follows.

1: Estimation (generate $\overline{\theta}_{k+1,i}$ and $\overline{P}_{k+1,i}$ based on $\widehat{\theta}_{k,i}$, $P_{k,i}$, $\varphi_{k,i}$, $s_{k+1,i}$ and $F_i(\cdot)$):

$$\overline{\theta}_{k+1,i} = \Pi_{\overline{P}^{-1}_{k+1,i}} \left\{\widehat{\theta}_{k,i} + a_{k,i} P_{k,i} \varphi_{k,i} \left(F_i(C - \varphi^{\mathrm{T}}_{k,i} \widehat{\theta}_{k,i}) - s_{k+1,i}\right)\right\}, \tag{3}$$

$$\overline{P}_{k+1,i} = P_{k,i} - a_{k,i} \underline{f} P_{k,i} \varphi_{k,i} \varphi^{\mathrm{T}}_{k,i} P_{k,i}, \tag{4}$$

$$a_{k,i} = \left(1 + \underline{f} \varphi^{\mathrm{T}}_{k,i} P_{k,i} \varphi_{k,i}\right)^{-1}, \tag{5}$$

2: Fusion (generate $\widehat{\theta}_{k+1,i}$ and $P^{-1}_{k+1,i}$ by a convex combination of $\overline{P}^{-1}_{k+1,j}$ and $\overline{\theta}_{k+1,i}$):

$$P^{-1}_{k+1,i} = \sum_{j\in\mathcal{N}_i} a_{ij} \overline{P}^{-1}_{k+1,j}, \tag{6}$$

$$\widehat{\theta}_{k+1,i} = P_{k+1,i} \sum_{j\in\mathcal{N}_i} a_{ij} \overline{P}^{-1}_{k+1,j} \overline{\theta}_{k+1,j}. \tag{7}$$

Remark 5 This algorithm is inspired by [27], whose algorithm is for the traditional single sensor case. They applied the specific projection operator $\Pi_Q\{\cdot\}$ into their algorithm, which overcame the difficulty of the quadratic item iterations in the Lyapunov function for the general projection operators. The proposed algorithm in this paper keeps this good characteristic as [27] by virtue of this specific projection operator, an ingenious fusion method as [8] and some inequalities on convex combinations of nonnegative definite matrices.

Let $\omega_{k+1,i} = F_i(C - \varphi^{\mathrm{T}}_{k,i}\theta) - s_{k+1,i}$ and $\Gamma_{k,i} = F_i(C - \varphi^{\mathrm{T}}_{k,i}\widehat{\theta}_{k,i}) - F_i(C - \varphi^{\mathrm{T}}_{k,i}\theta)$. Then, from (3) we have

$$\overline{\theta}_{k+1,i} = \Pi_{\overline{P}^{-1}_{k+1,i}} \left\{\widehat{\theta}_{k,i} + a_{k,i} P_{k,i} \varphi_{k,i} \left(\Gamma_{k,i} + \omega_{k+1,i}\right)\right\}. \tag{8}$$

Remark 6 One can verify that the sequence $\{\omega_{k,i}, \mathcal{F}_k\}$ is a martingale difference and $\sup_{k\geq 0} \mathbb{E}\left[|\omega_{k+1,i}|^{\beta} | \mathcal{F}_k\right] \leq 1$ for all $\beta > 0$ if Assumption 3 holds.

3 Main Results

For convenience of analysis, we introduce the following notations.

$$
\begin{aligned}
&Y_k := \mathrm{col}\{y_{k,1}, y_{k,2}, \cdots, y_{k,n}\}, && n\times 1\\
&S_k := \mathrm{col}\{s_{k,1}, s_{k,2}, \cdots, s_{k,n}\}, && n\times 1\\
&\varPhi_k := \mathrm{diag}\{\varphi_{k,1}, \varphi_{k,2}, \cdots, \varphi_{k,n}\}, && mn\times n\\
&\varGamma_k := \mathrm{col}\{\varGamma_{k,1}, \varGamma_{k,2}, \cdots, \varGamma_{k,n}\}, && n\times 1\\
&D_k := \mathrm{col}\{d_{k,1}, d_{k,2}, \cdots, d_{k,n}\}, && n\times 1\\
&W_k := \mathrm{col}\{\omega_{k,1}, \omega_{k,2}, \cdots, \omega_{k,n}\}, && n\times 1\\
&\varTheta := \mathrm{col}\{\theta, \theta, \cdots, \theta\}, && mn\times 1\\
&\widehat{\varTheta}_k := \mathrm{col}\left\{\widehat{\theta}_{k,1}, \widehat{\theta}_{k,2}, \cdots, \widehat{\theta}_{k,n}\right\}, && mn\times 1\\
&\widetilde{\varTheta}_k := \mathrm{col}\left\{\widetilde{\theta}_{k,1}, \widetilde{\theta}_{k,2}, \cdots, \widetilde{\theta}_{k,n}\right\}, \text{ where } \widetilde{\theta}_{k,i} := \widehat{\theta}_{k,i} - \theta, && mn\times 1\\
&\overline{\varTheta}_k := \mathrm{col}\left\{\overline{\theta}_{k,1}, \overline{\theta}_{k,2}, \cdots, \overline{\theta}_{k,n}\right\}, && mn\times 1\\
&\widetilde{\overline{\varTheta}}_k := \mathrm{col}\left\{\widetilde{\overline{\theta}}_{k,1}, \widetilde{\overline{\theta}}_{k,2}, \cdots, \widetilde{\overline{\theta}}_{k,n}\right\}, \text{ where } \widetilde{\overline{\theta}}_{k,i} := \overline{\theta}_{k,i} - \theta, && mn\times 1\\
&P_k := \mathrm{diag}\left\{P_{k,1}, P_{k,2}, \cdots, P_{k,n}\right\}, && mn\times mn\\
&\overline{P}_k := \mathrm{diag}\left\{\overline{P}_{k,1}, \overline{P}_{k,2}, \cdots, \overline{P}_{k,n}\right\}, && mn\times mn\\
&a_k := \mathrm{diag}\left\{a_{k,1}, a_{k,2}, \cdots, a_{k,n}\right\}, && n\times n\\
&b_k := a_k \otimes I_m, && mn\times mn\\
&\mathscr{A} := \mathcal{A} \otimes I_m, && mn\times mn
\end{aligned}
$$

where $\mathrm{col}\{\cdot\}$ denotes the vector stacked by the specified vectors, and $\mathrm{diag}\{\cdot\}$ denotes the block matrix formed in a diagonal manner of the corresponding vectors or matrices.

From the notations, the model (1) can be rewritten into the following matrix form,

$$
\begin{cases}
Y_{k+1} = \varPhi_k^{\mathrm{T}} \varTheta + D_{k+1},\\
S_{k+1} = \mathcal{S}(Y_{k+1}, C),
\end{cases}
$$

where $\mathcal{S}(Y_{k+1}, C) := \left[I_{\{y_{k+1,1}\le C\}}, I_{\{y_{k+1,2}\le C\}}, \cdots, I_{\{y_{k+1,n}\le C\}}\right]^{\mathrm{T}}$.

By (8) and the notations, the algorithm (3)–(7) can be written as

$$
\begin{cases}
b_k = a_k \otimes I_m,\\
\overline{\varTheta}_{k+1} = \boldsymbol{\varPi}_{\overline{P}_{k+1}^{-1}} \left\{\widehat{\varTheta}_k + b_k P_k \varPhi_k (\varGamma_k + W_{k+1})\right\},\\
\overline{P}_{k+1} = P_k - b_k \underline{f} P_k \varPhi_k \varPhi_k^{\mathrm{T}} P_k,\\
a_k = (1 + \underline{f}\, \varPhi_k^{\mathrm{T}} P_k \varPhi_k)^{-1},\\
\mathrm{vec}\left\{P_{k+1}^{-1}\right\} = \mathscr{A}\, \mathrm{vec}\{\overline{P}_{k+1}^{-1}\},\\
\widehat{\varTheta}_{k+1} = P_{k+1} \mathscr{A} \overline{P}_{k+1}^{-1} \overline{\varTheta}_{k+1},
\end{cases}
\tag{9}
$$

where $\mathrm{vec}\{\cdot\}$ denotes the operator that stacks the blocks of a block diagonal matrix on top of each other. And $\boldsymbol{\varPi}_H\{\cdot\}$ is defined as follows, for any positive definite matrix $H_i \in \mathbb{R}^{m\times m}$ and any vector $\zeta_i \in \mathbb{R}^m$, $i = 1, 2, \cdots, n$, define $\boldsymbol{\Pi}_H\{\zeta\} := (\Pi_{H_1}\{\zeta_1\}, \Pi_{H_2}\{\zeta_2\}, \cdots, \Pi_{H_n}\{\zeta_n\})^{\mathrm{T}}$, where $H = \operatorname{diag}\{H_1, H_2, \cdots, H_n\}$ and $\zeta = \operatorname{col}\{\zeta_1, \zeta_2, \cdots, \zeta_n\}$.

From (9) and $P_{k+1}\mathscr{A}\overline{P}_{k+1}^{-1}\vec{1} = \vec{1}$, we have

$$\begin{aligned}\widetilde{\Theta}_{k+1} = \widehat{\Theta}_{k+1} - \Theta &= P_{k+1}\mathscr{A}\overline{P}_{k+1}^{-1}\overline{\Theta}_{k+1} - \Theta \\ &= P_{k+1}\mathscr{A}\overline{P}_{k+1}^{-1}(\overline{\Theta}_{k+1} - \Theta) = P_{k+1}\mathscr{A}\overline{P}_{k+1}^{-1}\widetilde{\overline{\Theta}}_{k+1}.\end{aligned} \tag{10}$$

Before analyzing the convergence properties of the distributed recursive projection algorithm (3)–(7), we present a critical theorem about the relation between the estimation error and the regressors.

Theorem 1 *For the system* (1), *if Assumptions* 1–3 *hold, then the proposed algorithm* (3)–(7) *has the properties as* $t \to \infty$:

$$\begin{aligned}&\sum_{k=0}^{t}\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k = O(\log(r_t)), \ \ a.s., \\ &\widetilde{\Theta}_{t+1}^{\mathrm{T}}P_{t+1}^{-1}\widetilde{\Theta}_{t+1} = O(\log(r_t)), \ \ a.s.,\end{aligned}$$

where r_t *is defined as*

$$r_t = \lambda_{\max}(P_0) + \sum_{i=1}^{n}\sum_{k=0}^{t}\|\varphi_{k,i}\|^2. \tag{11}$$

The detailed proof of Theorem 1 is supplied in the next section.

Remark 7 From the proof of Lemma 1 and Theorem 1, we can see that the boundedness of the regressors and the unknown parameter (in Assumptions 1 and 2) is also a necessary condition to guarantee the convergence of the proposed algorithm in this paper. Similarly to accurate measurements, there is a positive definite item $\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k$ generated by the iteration of the covariance matrix P_k in the Lyapunov function analysis under binary-valued observations, which makes it difficult to estimate the bound of Lyapunov function. The boundedness of the inputs and the unknown parameter is exactly used to counteract the positive definite item so as to estimate the upper bound of the Lyapunov function.

Next, we will show the prediction ability of the distributed recursive projection algorithm based on Theorem 1. For any $i = 1, 2, \cdots, n$, the best prediction for the future output y_{k+1} in the mean square sense is given as the following form,

$$\mathbb{E}(y_{k+1,i}|\mathcal{F}_k) = \varphi_{k,i}^{\mathrm{T}}\theta + \mathbb{E}(d_{k+1,i}|\mathcal{F}_k).$$

Then, replacing θ by its estimate $\theta_{k,i}$ gives a natural adaptive predictor of $y_{k+1,i}$,

$$\widehat{y}_{k+1,i} = \varphi_{k,i}^{\mathrm{T}}\widehat{\theta}_{k,i} + \mathbb{E}(d_{k+1,i}|\mathcal{F}_k). \tag{12}$$

The difference between the best prediction and the adaptive prediction is referred to the regret $R_{k,i}$, which is denoted by,

$$R_{k,i} = [\mathbb{E}(y_{k+1,i}|\mathcal{F}_k) - \widehat{y}_{k+1,i}]^2 = (\varphi_{k,i}^{\mathrm{T}}\widetilde{\theta}_{k,i})^2. \tag{13}$$

The following theorem shows the upper bound of the accumulated regrets for the adaptive prediction (12).

Theorem 2 *Under Assumptions* 1–3*, the sample paths of the accumulated regrets have the following bound:*

$$\sum_{i=1}^{n}\sum_{k=0}^{t} R_{k,i} = O(\log(r_t)), \quad a.s. \tag{14}$$

The proof of Theorem 2 is given in Section 4.

Remark 8 From Theorem 2, it can be seen that the averaged regrets $\frac{1}{nt}\sum_{i=1}^{n}\sum_{k=0}^{t} R_{k,i}$ tends to zero as t goes to infinity under essentially no excitation condition on the regressors. Besides, there is also no need for independence, stationarity, or Gaussian assumption on system signals.

Assumption 5 (Cooperative excitation condition in [8]) The growth rate of $\log\left(\lambda_{\max}(P_t^{-1})\right)$ is much slower than that of $\lambda_{\min}(P_t^{-1})$, i.e.,

$$\lim_{t\to\infty}\frac{\log(r_t)}{\lambda_{\min}^{n,t}} = 0, \quad \text{a.s.},$$

where r_t is defined as (11) and

$$\lambda_{\min}^{n,t} := \lambda_{\min}\left(\sum_{i=1}^{n} P_{0,i}^{-1} + \sum_{i=1}^{n}\sum_{k=0}^{t-D(\mathcal{G})+1}\varphi_{k,i}\varphi_{k,i}^{\mathrm{T}}\right). \tag{15}$$

Remark 9 This cooperative excitation condition is the weakest possible data condition of the distributed LS estimates under accurate measurements to the best of our knowledge. Besause Assumption 5 could reduce to the well-known Lai-Wei excitation condition in [31], which is known to be the weakest possible data condition for the convergence of the classical LS estimates. Moreover, one can verify that the regressor condition in Assumption 2 and Assumption 5 is much weaker than the bounded persistence excitation conditions always used in the parameter estimation with binary-valued observations, such as $\|\varphi_l\| \leq M, \sum_{l=k}^{k+N-1}\varphi_l\varphi_l^{\mathrm{T}} \geq \varepsilon I_n$ in [20, 21] and $\|\varphi_l\| \leq M$, $\liminf_{k\to\infty}\frac{1}{k}\sum_{l=1}^{k}\varphi_l\varphi_l^{\mathrm{T}} > 0$ in [22, 23]. Thus, the bounded cooperative excitation condition (including Assumption 2 and Assumption 5) might be the weakest excitation condition of the distributed parameter estimation under binary-valued observations. Moreover, it is more general than independent signals and correlated non-stationary signals from feedback control systems, which makes the results in this paper can be applied to feedback control problems with binary observations.

The following theorem shows that the estimate given by the algorithm (3)–(7) converges to the true parameter under the bounded cooperative excitation condition.

Theorem 3 *For the system* (1)*, if Assumptions* 1–4 *are satisfied, then we have as* $t \to \infty$*,*

$$\left\|\widetilde{\Theta}_{t+1}\right\|^2 = O\left(\frac{\log(r_t)}{\lambda_{\min}^{n,t}}\right), \quad a.s.,$$

where r_t *and* $\lambda_{\min}^{n,t}$ *are defined in* (11) *and* (15)*, respectively.*

The detailed proof of Theorem 3 is given in Section 4.

Corollary 4 *Under Assumptions* 1–5*, the estimate* $\widehat{\theta}_{t,i}$ *given by the proposed algorithm* (3)–(7) *converges to the true parameter almost surely, i.e.,* $\lim_{t\to\infty}\widetilde{\theta}_{t,i}=0, i=1,2,\cdots,n$*, a.s.*

This conclusion can be directly given by Theorem 3 and Assumption 5.

Remark 10 From Theorem 3, we find that the proposed algorithm under binary-valued observations can reach same convergence rate as that in [8] under accurate observations under the same regressors condition (i.e., Assumption 2 and Assumption 5).

4 Proofs of the Main Results

Before establishing the convergence of the proposed algorithm (3)–(7), we give the following three lemmas to show the boundness of the estimates and two inequalities between P_k and $\overline{P}_k$.

Lemma 1 *With the algorithm* (3)–(7)*, the following inequality holds*

$$\left\|\widehat{\theta}_{k+1,i}\right\|\leq\overline{\theta},\quad i=1,2,\cdots,n,$$

for all $k=1,2,\cdots$ *under Assumption* 1*.*

Proof From Assumption 1, Definition 1 and (3), we have

$$\|\overline{\theta}_{k+1,i}\|\leq\overline{\theta}. \tag{16}$$

From (6), (7), and (16) we get

$$\begin{aligned}\left\|\widehat{\theta}_{k+1,i}\right\|&=\left\|P_{k+1,i}\sum_{j\in\mathcal{N}_i}a_{ij}\overline{P}_{k+1,j}^{-1}\overline{\theta}_{k+1,j}\right\|\\&\leq\left\|P_{k+1,i}\sum_{j\in\mathcal{N}_i}a_{ij}\overline{P}_{k+1,j}^{-1}\right\|\cdot\max_{1\leq j\leq n}\|\overline{\theta}_{k+1,j}\|\\&=\max_{1\leq j\leq n}\|\overline{\theta}_{k+1,j}\|\leq\overline{\theta}.\end{aligned}$$

From the definition of $\overline{P}_{k+1}$ and P_{k+1}, the following lemmas can be directly concluded similarly to Lemma 4.2, Lemma 4.3 and Lemma 4.4 in [8].

Lemma 2 *For any adjacency matrix* $\mathcal{A}=\{a_{ij}\}\in\mathbb{R}^{n\times n}$*, denote* $\mathscr{A}=\mathcal{A}\otimes I_m$*, and for all* $k=1,2,\cdots$*,*

$$\begin{aligned}&\mathscr{A}P_{k+1}\mathscr{A}\leq\overline{P}_{k+1},\\&\left|\overline{P}_{k+1}^{-1}\right|\leq\left|P_{k+1}^{-1}\right|,\end{aligned}$$

where $\overline{P}_{k+1}$ *and* P_{k+1} *are defined in* (9)*.*

Lemma 3 *With the distributed recursive projection algorithm* (9)*, the following equality holds*

$$\sum_{k=0}^{t}\lambda_{\max}(a_k\underline{f}\Phi_k^{\mathrm{T}}P_k\Phi_k)\leq\log\left|P_{t+1}^{-1}\right|-\log\left|P_0^{-1}\right|.$$

4.1 Proof of Theorem 1

Before proving Theorem 1, we need to show the following critical lemma.

Lemma 4 *If Assumptions* 1–3 *hold, then the proposed algorithm* (3)–(7) *satisfies the following relationship as* $t\to\infty$*:*

$$\widetilde{\Theta}_{t+1}^{\mathrm{T}}P_{t+1}^{-1}\widetilde{\Theta}_{t+1}+\left(\underline{f}+o(1)\right)\sum_{k=0}^{t}\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k\le\frac{2n}{\underline{f}}\log|P_{t+1}^{-1}|+o\left(\log|P_{t+1}^{-1}|\right)+O(1),\text{a.s.}$$

Proof From (9) and Remark 4, we get

$$\begin{aligned}
&\widetilde{\overline{\Theta}}_{k+1}^{\mathrm{T}}\overline{P}_{k+1}^{-1}\widetilde{\overline{\Theta}}_{k+1}\\
&=\left(\boldsymbol{\Pi}_{\overline{P}_{k+1}^{-1}}\left\{\widehat{\Theta}_k+b_kP_k\Phi_k(\Gamma_k+W_{k+1})\right\}-\Theta\right)^{\mathrm{T}}\overline{P}_{k+1}^{-1}\\
&\quad\cdot\left(\boldsymbol{\Pi}_{\overline{P}_{k+1}^{-1}}\left\{\widehat{\Theta}_k+b_kP_k\Phi_k(\Gamma_k+W_{k+1})\right\}-\Theta\right)\\
&=\sum_{i=1}^{n}\left\|\Pi_{\overline{P}_{k+1,i}^{-1}}\left\{\widehat{\theta}_{k,i}+a_{k,i}P_{k,i}\varphi_{k,i}\left(\Gamma_{k,i}+\omega_{k+1,i}\right)\right\}-\theta\right\|_{\overline{P}_{k+1,i}^{-1}}^2\\
&\le\sum_{i=1}^{n}\left\|\widetilde{\theta}_{k,i}+a_{k,i}P_{k,i}\varphi_{k,i}\left(\Gamma_{k,i}+\omega_{k+1,i}\right)\right\|_{\overline{P}_{k+1,i}^{-1}}^2\\
&\le\sum_{i=1}^{n}\left(\widetilde{\theta}_{k,i}+a_{k,i}P_{k,i}\varphi_{k,i}\left(\Gamma_{k,i}+\omega_{k+1,i}\right)\right)^{\mathrm{T}}\overline{P}_{k+1,i}^{-1}\\
&\quad\cdot\left(\widetilde{\theta}_{k,i}+a_{k,i}P_{k,i}\varphi_{k,i}\left(\Gamma_{k,i}+\omega_{k+1,i}\right)\right)\\
&=\left(\widetilde{\Theta}_k+b_kP_k\Phi_k(\Gamma_k+W_{k+1})\right)^{\mathrm{T}}\overline{P}_{k+1}^{-1}\left(\widetilde{\Theta}_k+b_kP_k\Phi_k(\Gamma_k+W_{k+1})\right).
\end{aligned}\tag{17}$$

Then, by (10), (17), and Lemma 2, the stochastic Lyapunov function $V_k=\widetilde{\Theta}_k^{\mathrm{T}}P_k^{-1}\widetilde{\Theta}_k$ satisfies

$$\begin{aligned}
V_{k+1}&=\widetilde{\Theta}_{k+1}^{\mathrm{T}}P_{k+1}^{-1}\widetilde{\Theta}_{k+1}\\
&=\left(P_{k+1}\mathscr{A}\overline{P}_{k+1}^{-1}\widetilde{\overline{\Theta}}_{k+1}\right)^{\mathrm{T}}P_{k+1}^{-1}\left(P_{k+1}\mathscr{A}\overline{P}_{k+1}^{-1}\widetilde{\overline{\Theta}}_{k+1}\right)\\
&=\widetilde{\overline{\Theta}}_{k+1}^{\mathrm{T}}\overline{P}_{k+1}^{-1}\mathscr{A}P_{k+1}\mathscr{A}\overline{P}_{k+1}^{-1}\widetilde{\overline{\Theta}}_{k+1}\\
&\le\widetilde{\overline{\Theta}}_{k+1}^{\mathrm{T}}\overline{P}_{k+1}^{-1}\widetilde{\overline{\Theta}}_{k+1}\\
&\le\left(\widetilde{\Theta}_k+b_kP_k\Phi_k(\Gamma_k+W_{k+1})\right)^{\mathrm{T}}\overline{P}_{k+1}^{-1}\left(\widetilde{\Theta}_k+b_kP_k\Phi_k(\Gamma_k+W_{k+1})\right)\\
&=\widetilde{\Theta}_k^{\mathrm{T}}\overline{P}_{k+1}^{-1}\widetilde{\Theta}_k+2\widetilde{\Theta}_k^{\mathrm{T}}\overline{P}_{k+1}^{-1}b_kP_k\Phi_k\Gamma_k+2\widetilde{\Theta}_k^{\mathrm{T}}\overline{P}_{k+1}^{-1}b_kP_k\Phi_kW_{k+1}\\
&\quad+\Gamma_k^{\mathrm{T}}\Phi_k^{\mathrm{T}}P_kb_k\overline{P}_{k+1}^{-1}b_kP_k\Phi_k\Gamma_k+W_{k+1}^{\mathrm{T}}\Phi_k^{\mathrm{T}}P_kb_k\overline{P}_{k+1}^{-1}b_kP_k\Phi_kW_{k+1}\\
&\quad+2\Gamma_k^{\mathrm{T}}\Phi_k^{\mathrm{T}}P_kb_k\overline{P}_{k+1}^{-1}b_kP_k\Phi_kW_{k+1}.
\end{aligned}\tag{18}$$

By (9), we have

$$\overline{P}_{k+1}^{-1}=P_k^{-1}+\underline{f}\Phi_k\Phi_k^{\mathrm{T}}.\tag{19}$$

So, from (19), the first term in the right side of the equality (18) can be rewritten as

$$\widetilde{\Theta}_k^{\mathrm{T}}\overline{P}_{k+1}^{-1}\widetilde{\Theta}_k=\widetilde{\Theta}_k^{\mathrm{T}}P_k^{-1}\widetilde{\Theta}_k+\underline{f}\,\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k=V_k+\underline{f}\,\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k. \tag{20}$$

Moreover, from the block diagonal property of a_k, b_k, P_k, and Φ_k, we get

$$b_kP_k=P_kb_k,\quad \Phi_k^{\mathrm{T}}b_k=a_k\Phi_k^{\mathrm{T}},\quad b_k\Phi_k=\Phi_ka_k. \tag{21}$$

From differential mean value theorem, there exists $\xi_{k,i}\in(C-\varphi_{k,i}^{\mathrm{T}}\widehat{\theta}_{k,i},C-\varphi_{k,i}^{\mathrm{T}}\theta)$ or $(C-\varphi_{k,i}^{\mathrm{T}}\theta,C-\varphi_{k,i}^{\mathrm{T}}\widehat{\theta}_{k,i})$ such that

$$\Gamma_{k,i}=F_i(C-\varphi_{k,i}^{\mathrm{T}}\widehat{\theta}_{k,i})-F_i(C-\varphi_{k,i}^{\mathrm{T}}\theta)=-f_i(\xi_{k,i})\varphi_{k,i}^{\mathrm{T}}\widetilde{\theta}_{k,i}. \tag{22}$$

Then, from Lemma 1 and Assumption 2, we get $\xi_{k,i}\in(C-\overline{\theta}\overline{\varphi},C+\overline{\theta}\overline{\varphi})$. From Assumption 3, we have

$$f_i(\xi_{k,i})\geq\underline{f}. \tag{23}$$

By (19), (21)–(23), and $a_k(I_n+\underline{f}\,\Phi_k^{\mathrm{T}}P_k\Phi_k)=I_n$, we can estimate the second term in the right side of the equality (18) as follows:

$$\begin{aligned}2\widetilde{\Theta}_k^{\mathrm{T}}\overline{P}_{k+1}^{-1}b_kP_k\Phi_k\Gamma_k=&2\widetilde{\Theta}_k^{\mathrm{T}}\left(P_k^{-1}+\underline{f}\,\Phi_k\Phi_k^{\mathrm{T}}\right)b_kP_k\Phi_k\Gamma_k\\=&2\widetilde{\Theta}_k^{\mathrm{T}}\Phi_ka_k\left(I_n+\underline{f}\,\Phi_k^{\mathrm{T}}P_k\Phi_k\right)\Gamma_k\\=&2\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Gamma_k\\=&2\sum_{i=1}^n\widetilde{\theta}_{k,i}^{\mathrm{T}}\varphi_{k,i}\Gamma_{k,i}\\=&-2\sum_{i=1}^nf_i(\xi_{k,i})\widetilde{\theta}_{k,i}^{\mathrm{T}}\varphi_{k,i}\varphi_{k,i}^{\mathrm{T}}\widetilde{\theta}_{k,i}\\\leq&-2\underline{f}\sum_{i=1}^n\widetilde{\theta}_{k,i}^{\mathrm{T}}\varphi_{k,i}\varphi_{k,i}^{\mathrm{T}}\widetilde{\theta}_{k,i}\\=&-2\underline{f}\,\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k.\end{aligned} \tag{24}$$

Similarly to (24), the third term in the right side of the equality (18) can be estimated as,

$$\begin{aligned}2\widetilde{\Theta}_k^{\mathrm{T}}\overline{P}_{k+1}^{-1}b_kP_k\Phi_kW_{k+1}=&2\widetilde{\Theta}_k^{\mathrm{T}}\left(P_k^{-1}+\underline{f}\,\Phi_k\Phi_k^{\mathrm{T}}\right)b_kP_k\Phi_kW_{k+1}\\=&2\widetilde{\Theta}_k^{\mathrm{T}}\Phi_ka_k\left(I_n+\underline{f}\,\Phi_k^{\mathrm{T}}P_k\Phi_k\right)W_{k+1}\\=&2\widetilde{\Theta}_k^{\mathrm{T}}\Phi_kW_{k+1}.\end{aligned} \tag{25}$$

From (19), (21), and $I_n=a_k+a_k\underline{f}\,\Phi_k^{\mathrm{T}}P_k\Phi_k$, we have

$$\begin{aligned}\Phi_k^{\mathrm{T}}P_kb_k\overline{P}_{k+1}^{-1}b_kP_k\Phi_k=&\Phi_k^{\mathrm{T}}P_kb_k(P_k^{-1}+\underline{f}\,\Phi_k\Phi_k^{\mathrm{T}})b_kP_k\Phi_k\\=&\Phi_k^{\mathrm{T}}P_kb_k^2\Phi_k+\underline{f}\,\Phi_k^{\mathrm{T}}P_kb_k\Phi_k\Phi_k^{\mathrm{T}}b_kP_k\Phi_k\\=&a_k^2\Phi_k^{\mathrm{T}}P_k\Phi_k+a_k\underline{f}\,\Phi_k^{\mathrm{T}}P_k\Phi_ka_k\Phi_k^{\mathrm{T}}P_k\Phi_k\\=&a_k^2\Phi_k^{\mathrm{T}}P_k\Phi_k+(I_n-a_k)a_k\Phi_k^{\mathrm{T}}P_k\Phi_k\\=&a_k\Phi_k^{\mathrm{T}}P_k\Phi_k.\end{aligned} \tag{26}$$

Then, from (26), Assumption 3 and $|\Gamma_{k,i}| \leq 1$, the fourth term in the right side of the equality (18) can be estimated as,

$$\begin{aligned}\Gamma_k^{\mathrm{T}} \Phi_k^{\mathrm{T}} P_k b_k \overline{P}_{k+1}^{-1} b_k P_k \Phi_k \Gamma_k =& \Gamma_k^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k \Gamma_k \\ =& \sum_{i=1}^{n} a_{k,i} \varphi_{k,i}^{\mathrm{T}} P_{k,i} \varphi_{k,i} \Gamma_{k,i}^2 \\ \leq& n \|a_k \Phi_k^{\mathrm{T}} P_k \Phi_k\| \\ =& n\lambda_{\max}(a_k \Phi_k^{\mathrm{T}} P_k \Phi_k).\end{aligned} \tag{27}$$

From (26), we can rewrite the fifth and sixth terms in the right side of the equality (18) as

$$W_{k+1}^{\mathrm{T}} \Phi_k^{\mathrm{T}} P_k b_k \overline{P}_{k+1}^{-1} b_k P_k \Phi_k W_{k+1} = W_{k+1}^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1} \tag{28}$$

and

$$2\Gamma_k^{\mathrm{T}} \Phi_k^{\mathrm{T}} P_k b_k \overline{P}_{k+1}^{-1} b_k P_k \Phi_k W_{k+1} = 2\Gamma_k^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1}. \tag{29}$$

Taking (20), (24)–(25) and (27)–(29) into (18) gives

$$\begin{aligned}V_{k+1} \leq& V_k - \underline{f}\, \widetilde{\Theta}_k^{\mathrm{T}} \Phi_k \Phi_k^{\mathrm{T}} \widetilde{\Theta}_k + n\lambda_{\max}(a_k \Phi_k^{\mathrm{T}} P_k \Phi_k) + 2\widetilde{\Theta}_k^{\mathrm{T}} \Phi_k a_k W_{k+1} \\ &+ 2\Gamma_k^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1} + W_{k+1}^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1}.\end{aligned} \tag{30}$$

For (30), summing from $k=0$ to t yields

$$\begin{aligned}V_{t+1} + \underline{f} \sum_{k=0}^{t} \widetilde{\Theta}_k^{\mathrm{T}} \Phi_k \Phi_k^{\mathrm{T}} \widetilde{\Theta}_k \leq& V_0 + n\sum_{k=0}^{t} \lambda_{\max}(a_k \Phi_k^{\mathrm{T}} P_k \Phi_k) + 2\sum_{k=0}^{t} \widetilde{\Theta}_k^{\mathrm{T}} \Phi_k W_{k+1} \\ &+ 2\sum_{k=0}^{t} \Gamma_k^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1} + \sum_{k=0}^{t} W_{k+1}^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1}.\end{aligned} \tag{31}$$

Next, we estimate the third and fourth terms in the side of the equality (31). By Remark 6, $\widetilde{\Theta}_k^{\mathrm{T}} \Phi_k \in \mathcal{F}_k$, $a_k \underline{f} \Phi_k^{\mathrm{T}} P_k \Phi_k = I_n - a_k$ and $0 < a_k \leq I_n$, we can get the following estimates for any $\delta > 0$ using the martingale estimation theorem (Theorem 2.8 in [32]):

$$\sum_{k=0}^{t} \widetilde{\Theta}_k^{\mathrm{T}} \Phi_k W_{k+1} = O\left(\left[\sum_{k=0}^{t} \left\|\widetilde{\Theta}_k^{\mathrm{T}} \Phi_k \Phi_k^{\mathrm{T}} \widetilde{\Theta}_k\right\|\right]^{\frac{1}{2}+\delta}\right), \quad \text{a.s.} \tag{32}$$

and

$$\begin{aligned}\sum_{k=0}^{t} \Gamma_k^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1} =& O\left(\left[\sum_{k=0}^{t} \left\|\Gamma_k^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k \Phi_k^{\mathrm{T}} P_k \Phi_k a_k \Gamma_k\right\|\right]^{\frac{1}{2}+\delta}\right) \\ =& O\left(\left[\sum_{k=0}^{t} n \left\|a_k \Phi_k^{\mathrm{T}} P_k \Phi_k\right\|^2\right]^{\frac{1}{2}+\delta}\right)\end{aligned}$$

$$
\begin{aligned}
&\le O\left(\left[\sum_{k=0}^{t}\frac{n}{\underline{f}}\left\|a_k\Phi_k^{\mathrm{T}}P_k\Phi_k\right\|\right]^{\frac{1}{2}+\delta}\right)\\
&=O\left(\left[\sum_{k=0}^{t}\left\|a_k\Phi_k^{\mathrm{T}}P_k\Phi_k\right\|\right]^{\frac{1}{2}+\delta}\right),\quad \text{a.s.}
\end{aligned}
\tag{33}
$$

By (32), (33), and taking $0<\delta<\frac{1}{2}$, we have

$$
\sum_{k=0}^{t}\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k W_{k+1}=o\left(\sum_{k=0}^{t}\left\|\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k\right\|\right)+O(1),\quad \text{a.s.}
\tag{34}
$$

and

$$
\sum_{k=0}^{t}\Gamma_k^{\mathrm{T}}a_k\Phi_k^{\mathrm{T}}P_k\Phi_k W_{k+1}=o\left(\sum_{k=0}^{t}\left\|a_k\Phi_k^{\mathrm{T}}P_k\Phi_k\right\|\right)+O(1),\quad \text{a.s.}
\tag{35}
$$

We now proceed to estimate the last term in (31).

$$
W_{k+1}^{\mathrm{T}}a_k\Phi_k^{\mathrm{T}}P_k\Phi_k W_{k+1}\le\|a_k\Phi_k^{\mathrm{T}}P_k\Phi_k\|\cdot\|W_{k+1}\|^2=\lambda_{\max}(a_k\Phi_k^{\mathrm{T}}P_k\Phi_k)\cdot\left\{\sum_{i=1}^{n}\omega_{k+1,i}^2\right\}.
\tag{36}
$$

Denote a martingale difference sequence $\varpi_{k+1}=\sum_{i=1}^{n}\omega_{k+1,i}^2-\mathbb{E}[\sum_{i=1}^{n}\omega_{k+1,i}^2|\mathcal{F}_k]$. By Remark 6, C_r-inequality (Theorem 1.2.12 in [33]) and Lyapunov inequality (below Theorem 1.4 in [32]), it is easy to see that for any $\alpha\in(2,\min(\beta,4))$,

$$
\begin{aligned}
\sup_k\mathbb{E}[\varpi_{k+1}^{\frac{\alpha}{2}}|\mathcal{F}_k]&=\sup_k\mathbb{E}\left[\left(\sum_{i=1}^{n}\omega_{k+1,i}^2-\mathbb{E}\left[\sum_{i=1}^{n}\omega_{k+1,i}^2|\mathcal{F}_k\right]\right)^{\frac{\alpha}{2}}\middle|\mathcal{F}_k\right]\\
&\le 4\sup_k\mathbb{E}\left[\sum_{i=1}^{n}|\omega_{k+1,i}|^{\alpha}\middle|\mathcal{F}_k\right]<\infty,\quad \text{a.s.}
\end{aligned}
$$

Then, by the martingale estimation theorem (Theorem 2.8 in [32]), we have for any $\eta>0$,

$$
\begin{aligned}
&\sum_{k=0}^{t}\lambda_{\max}(a_k\Phi_k^{\mathrm{T}}P_k\Phi_k)\left(\sum_{i=1}^{n}(\omega_{k+1,i})^2-\mathbb{E}\left[\sum_{i=1}^{n}(\omega_{k+1,i})^2\middle|\mathcal{F}_k\right]\right)\\
&=O\left(S_t\left(\frac{\alpha}{2}\right)\left[\log\left(S_t\left(\frac{\alpha}{2}\right)+e\right)\right]^{\frac{2}{\alpha}+\eta}\right),\quad \text{a.s.,}
\end{aligned}
\tag{37}
$$

where

$$
S_t\left(\frac{\alpha}{2}\right)=\left[\sum_{k=0}^{t}(\lambda_{\max}(a_k\Phi_k^{\mathrm{T}}P_k\Phi_k))^{\frac{\alpha}{2}}\right]^{\frac{2}{\alpha}}=\frac{1}{\underline{f}}\left[\sum_{k=0}^{t}(\lambda_{\max}(a_k\underline{f}\,\Phi_k^{\mathrm{T}}P_k\Phi_k))^{\frac{\alpha}{2}}\right]^{\frac{2}{\alpha}}.
$$

Since $\frac{\alpha}{2}>1$, $a_k\underline{f}\,\Phi_k^{\mathrm{T}}P_k\Phi_k\le I_n$ and Lemma 3 we have

$$
S_t\left(\frac{\alpha}{2}\right)=O(1)+o\left(\log|P_{t+1}^{-1}|\right).
$$

From (36)–(37), Remark 6 and Lemma 3, we get

$$\begin{aligned}
& W_{k+1}^{\mathrm{T}} a_k \Phi_k^{\mathrm{T}} P_k \Phi_k W_{k+1} \\
\leq & \lambda_{\max}(a_k \Phi_k^{\mathrm{T}} P_k \Phi_k) \left(\mathbb{E}\left[\sum_{i=1}^{n} \omega_{k+1,i}^2 \middle| \mathcal{F}_k \right] + \sum_{i=1}^{n} \omega_{k+1,i}^2 - \mathbb{E}\left[\sum_{i=1}^{n} \omega_{k+1,i}^2 \middle| \mathcal{F}_k \right] \right) \\
= & n \sum_{k=0}^{t} \lambda_{\max}(a_k \Phi_k^{\mathrm{T}} P_k \Phi_k) + O(1) + o\left(\log |P_{t+1}^{-1}|\right) \\
= & \frac{n}{\underline{f}} \log |P_{t+1}^{-1}| + o(\log |P_{t+1}^{-1}|) + O(1), \quad \text{a.s.}
\end{aligned} \tag{38}$$

Finally, from Lemma 3 , taking (34), (35), and (38) into (31) gives

$$\widetilde{\Theta}_{t+1}^{\mathrm{T}} P_{t+1}^{-1} \widetilde{\Theta}_{t+1} + \left(\underline{f} + o(1)\right) \sum_{k=0}^{t} \widetilde{\Theta}_k^{\mathrm{T}} \Phi_k \Phi_k^{\mathrm{T}} \widetilde{\Theta}_k \leq \frac{2n}{\underline{f}} \log |P_{t+1}^{-1}| + o\left(\log |P_{t+1}^{-1}|\right) + O(1), \quad \text{a.s.}$$

This completes the proof. ∎

Based on Lemma 4, we give the proof of Theorem 1 as follows.

Proof of Theorem 1

Proof From (6) we get that for any $t \geq 0$,

$$P_{t+1,i}^{-1} = \sum_{j=1}^{n} a_{ij} \overline{P}_{t+1,j}^{-1} = \sum_{j=1}^{n} a_{ij} \left[P_{t,j}^{-1} + \underline{f} \varphi_{t,j} \varphi_{t,j}^{\mathrm{T}} \right].$$

Therefore, we have

$$\begin{aligned}
\max_{1 \leq i \leq n} \lambda_{\max}\left(P_{t+1,i}^{-1}\right) & \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} a_{ij} \left[\lambda_{\max}\left(P_{t,j}^{-1}\right) + \underline{f} \lambda_{\max}(\varphi_{t,j} \varphi_{t,j}^{\mathrm{T}}) \right] \\
& \leq \max_{1 \leq i \leq n} \lambda_{\max}\left(P_{t,i}^{-1}\right) \sum_{j=1}^{n} a_{ij} + \underline{f} \sum_{j=1}^{n} \lambda_{\max}(\varphi_{t,j} \varphi_{t,j}^{\mathrm{T}}) \\
& \leq \max_{1 \leq i \leq n} \lambda_{\max}\left(P_{t,i}^{-1}\right) + \underline{f} \sum_{j=1}^{n} \|\varphi_{t,j}\|^2 \\
& \leq \max_{1 \leq i \leq n} \lambda_{\max}\left(P_{t-1,i}^{-1}\right) + \underline{f} \sum_{j=1}^{n} \sum_{k=t-1}^{t} \|\varphi_{k,j}\|^2 \\
& \leq \cdots \\
& \leq \max_{1 \leq i \leq n} \lambda_{\max}\left(P_{0,i}^{-1}\right) + \underline{f} \sum_{j=1}^{n} \sum_{k=0}^{t} \|\varphi_{k,j}\|^2 \\
& = \lambda_{\max}\left(P_0^{-1}\right) + \underline{f} \sum_{j=1}^{n} \sum_{k=0}^{t} \|\varphi_{k,j}\|^2.
\end{aligned} \tag{39}$$

From (39) and the connection between determinant and eigenvalues of the matrix, one can conclude that

$$\log(|P_{t+1}^{-1}|) \leq mn \log\left(\max_{1\leq i\leq n} \lambda_{\max}\left(P_{t+1,i}^{-1}\right)\right) \leq mn\log(r_t) + O(1). \tag{40}$$

Consequently, Theorem 1 follows from (40) and Lemma 4 immediately. ∎

4.2 Proof of Theorem 2

Proof From (13), we have

$$\sum_{i=1}^{n}\sum_{k=0}^{t} R_{k,i} = \sum_{i=1}^{n}\sum_{k=0}^{t}(\varphi_{k,i}^{\mathrm{T}}\widetilde{\theta}_{k,i})^2 = \sum_{k=0}^{t}\widetilde{\Theta}_k^{\mathrm{T}}\Phi_k\Phi_k^{\mathrm{T}}\widetilde{\Theta}_k. \tag{41}$$

Then, substituting (41) into Theorem 1 gives (14) immediately. ∎

4.3 Proof of Theorem 3

Proof From Assumption 4 and Remark 3, we have $a_{ij}^{(D_{\mathcal{G}})} \geq \underline{a}$, where $\underline{a} = \min_{i,j\in\mathcal{V}} a_{ij}^{(D_{\mathcal{G}})} > 0$, $D_{\mathcal{G}}$ is diameter of the graph $\mathcal{G}$. And, one can also prove that $a_{ij}^{(d)} \geq \underline{a}$ by induction when $k \geq D_{\mathcal{G}}$, where $\mathcal{A}^d = \left\{a_{ij}^{(d)}\right\}$. By (9) and (19), it can be seen

$$\begin{aligned}
\mathrm{vec}\{P_{t+1}^{-1}\} &= \mathscr{A}\,\mathrm{vec}\{\overline{P}_{t+1}^{-1}\}\\
&= \mathscr{A}\,\mathrm{vec}\{P_t^{-1}\} + \underline{f}\mathscr{A}\,\mathrm{vec}\{\Phi_t\Phi_t^{\mathrm{T}}\}\\
&= \mathscr{A}^2\,\mathrm{vec}\{P_{t-1}^{-1}\} + \underline{f}\mathscr{A}^2\,\mathrm{vec}\{\Phi_{t-1}\Phi_{t-1}^{\mathrm{T}}\} + \underline{f}\mathscr{A}\,\mathrm{vec}\{\Phi_t\Phi_t^{\mathrm{T}}\}\\
&= \cdots\\
&= \mathscr{A}^{t+1}\,\mathrm{vec}\{P_0^{-1}\} + \underline{f}\sum_{k=0}^{t}\mathscr{A}^{t-k+1}\,\mathrm{vec}\{\Phi_k\Phi_k^{\mathrm{T}}\},
\end{aligned}$$

which implies

$$\begin{aligned}
P_{t+1,i}^{-1} &= \sum_{j=1}^{n} a_{ij}^{(t+1)}P_{0,j}^{-1} + \sum_{j=1}^{n}\sum_{k=0}^{t} a_{ij}^{(t-k+1)}\underline{f}\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}\\
&\geq \underline{a}\sum_{j=1}^{n}P_{0,j}^{-1} + \underline{af}\sum_{j=1}^{n}\sum_{k=0}^{t-D_{\mathcal{G}}+1}\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}.
\end{aligned}$$

Hence, we have

$$\lambda_{\min}\left(P_{t+1}^{-1}\right) \geq \min\left\{\underline{a}, \underline{af}\right\}\cdot\lambda_{\min}\left(\sum_{j=1}^{n}P_{0,j}^{-1} + \sum_{j=1}^{n}\sum_{k=0}^{t-D_{\mathcal{G}}+1}\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}\right). \tag{42}$$

Noticing that

$$\left\|\widetilde{\Theta}_{t+1}\right\|^2 \leq \widetilde{\Theta}_{t+1}^{\mathrm{T}}\frac{P_{t+1}^{-1}}{\lambda_{\min}\left(P_{t+1}^{-1}\right)}\widetilde{\Theta}_{t+1},$$

we have

$$\left\|\widetilde{\Theta}_{t+1}\right\|^2 = O\left(\frac{\log(r_t)}{\lambda_{\min}^{n,t}}\right), \text{ a.s.,}$$

by (42) and Theorem 1. The proof is completed. ∎

5 Simulation

In this section, we illustrate the cooperative effect of the sensors by Example 1, i.e., the sensors in the network can achieve the estimation task that cannot be realized by any individual sensor through exchanging information between sensors. Besides, we also show the property of the averaged regret.

Example 1 The network is composed of $n = 3$ binary sensors whose dynamics obey the equation (1) with $m = 3$, and the threshold of binary sensors is $C = 0$. The adjacency matrix is taken as

$$\mathcal{A} = \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/2 & 1/6 \\ 0 & 1/6 & 5/6 \end{pmatrix}.$$

Set the unknown parameter as $\theta = [-1, 1, -2]^{\mathrm{T}}$, and the prior information of the unknown parameter is $\theta \in \Omega = [-2, 2] \times [-3, 3] \times [-4, 4]$. The system noises $\{d_{k,i}\}$, $i = 1, 2, 3$, in (1) are i.i.d. with $d_{k,i} \sim N(0, 4^2)$ (Gaussian distribution with zero mean and variance 4^2). Let the regressors $\varphi_{k,i}$ $(i = 1, 2, 3)$ be generated as follows:

$$\varphi_{k,1} = \begin{pmatrix} 1 - 1/3^k \\ 0 \\ 0 \end{pmatrix}, \quad \varphi_{k,2} = \begin{pmatrix} 0 \\ -1 + 1/4^k \\ 0 \end{pmatrix}, \quad \varphi_{k,3} = \begin{pmatrix} 0 \\ 0 \\ 1 - 1/2^k \end{pmatrix}, \quad k \geq 1.$$

One can verify that the regressors $\varphi_{k,i}$ $(i = 1, 2, 3)$ of the three sensors can cooperate to satisfy Assumptions 2 and 5. Besides, we take $\underline{f} = 0.04$ from $\overline{\varphi} = 1$, $\overline{\theta} = \sqrt{29}$ and $d_{k,i} \sim N(0, 4^2)$. Then we repeat the simulation for 100 times with the same initial values, where $P_{0,1} = P_{0,2} = P_{0,3} = I_3$. The mean square errors (MSE) of the three sensors (averaged over 100 runs) are shown in Figure 1. From it, we learn that the proposed algorithm (3)–(7) can achieve the distributed parameter estimation with binary-valued observations.

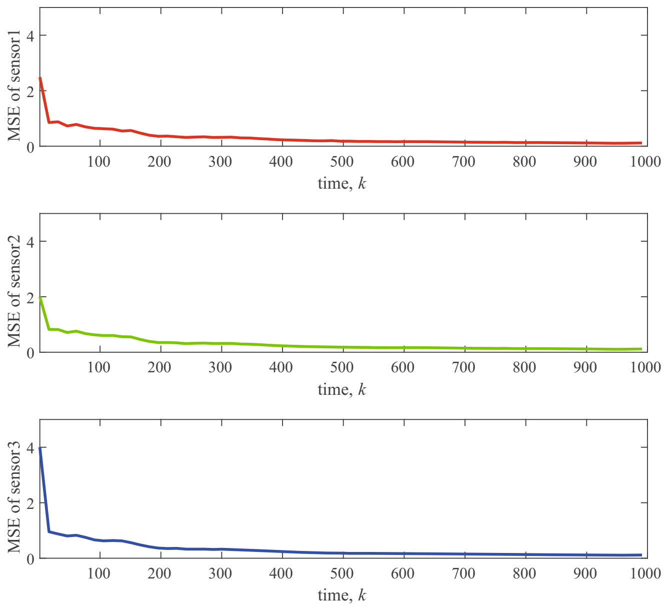

Figure 1 MSEs of the distributed recursive projection algorithm

Moreover, the averaged regret $\frac{1}{3k}\sum_{i=1}^{3}\sum_{l=0}^{k}R_{l,i}$ is showed in Figure 2, which implies that the averaged regrets tends to zero as k goes to infinity.

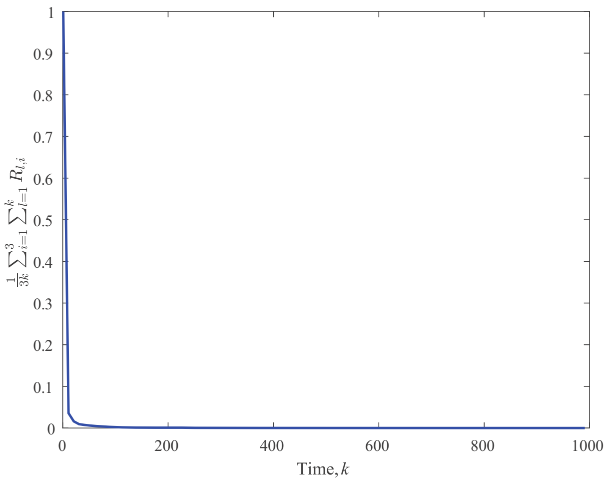

Figure 2 The averaged regrets

Furthermore, from $\lambda_{\max}(P_{0,i})+\sum_{k=0}^{t}\|\varphi_{k,i}\|^2\leq t+1$ and $\lambda_{\min}\left(P_{0,i}^{-1}+\sum_{k=0}^{t}\varphi_{k,i}\varphi_{k,i}^{\rm T}\right)=0$ for all $i=1,2,3$, we can see that none of the regressors $\varphi_{k,i}$, $i=1,2,3$, of the three individual sensors can satisfy the excitation condition, i.e.,

$$\lim_{t\to\infty}\frac{\log\left(\lambda_{\max}(P_{0,i})+\sum_{k=0}^{t}\|\varphi_{k,i}\|^2\right)}{\lambda_{\min}\left(P_{0,i}^{-1}+\sum_{k=0}^{t}\varphi_{k,i}\varphi_{k,i}^{\rm T}\right)}=0,\quad \text{a.s.}$$

Figure 3 shows the trajectory of the MSE of three sensors using the non-cooperative algorithm, i.e., the algorithm (3)–(7) with $\mathcal{A} = I_n$. Comparing Figure 1 and Figure 3, we learn that the MSE of each sensor using the non-cooperative algorithm does not converge to zero while the estimate given by the distributed recursive projection algorithm (3)–(7) converges to the true parameter, which shows the joint effect of the sensors.

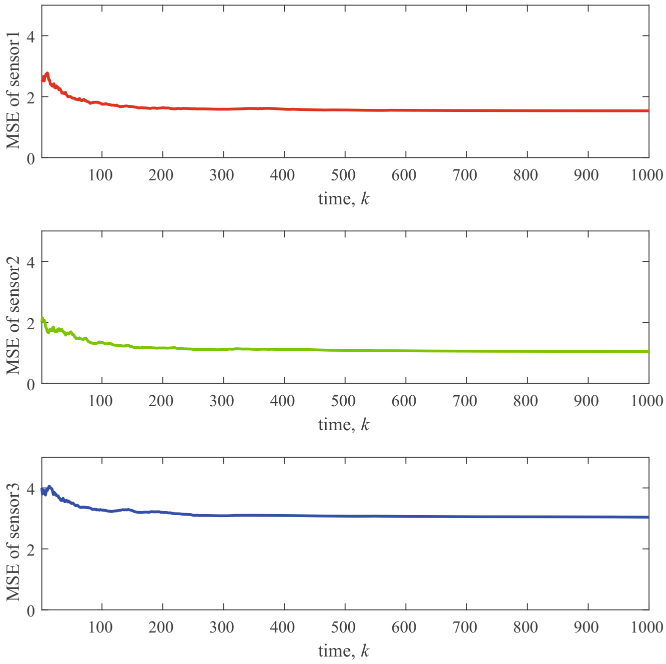

Figure 3 MSEs of the non-cooperative algorithm

6 Concluding Remarks

This paper has investigated a distributed recursive projection algorithm to estimate cooperatively the unknown parameter in the sensor network under binary-valued observations. For the adaptive predictor based on the proposed algorithm, we establish the upper bound of the accumulated regrets with no excitation condition on the regressors, which implies the averaged regret tends to zero as t increases to infinity. Then, we introduce the weakest cooperative excitation condition with binary-valued observations to the best of our knowledge, under which the almost sure convergence of the proposed algorithm are proved. Besides, we reveal the operative effect of the sensors through a numerical example. Furthermore, the results in this paper are obtained without relying on the independent or periodic assumptions on the regressors compared with most of the existing results, which makes it possible to apply our results to the feedback control systems with binary observations.

There are still lots of interesting problems for further research. For example, to consider the distributed algorithm with general projection (see [20, 22]) instead of this special projection, to consider the distributed optimal algorithm that can reach Cramér-Rao lower bound based on [23], to consider the distributed estimation problem for more general systems such as ARMAX systems in [34, 35], etc.

Acknowledgements Sincerely thank Miss Die Gan for her valuable discussions.

References

[1] Sayed A H, Tu S, Chen J, et al., Diffusion strategies for adaptation and learning over networks: An examination of distributed strategies and network behavior, *IEEE Signal Processing Magazine*, 2013, **30**(3): 155–171.

[2] Taj M and Cavallaro A, Distributed and decentralized multicamera tracking, *IEEE Signal Processing Magazine*, 2011, **28**(3): 46–58.

[3] Takahashi N, Yamada I, and Sayed A H, Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis, *IEEE Transactions on Signal Processing*, 2010, **58**(9): 4795–4810.

[4] Zhang Q and Zhang J F, Distributed parameter estimation over unreliable networks with markovian switching topologies, *IEEE Transactions on Automatic Control*, 2012, **57**(10): 2545–2560.

[5] Sun Y, Fu M, and Zhang H, Performance comparison of distributed state estimation algorithms for power systems, *Journal of Systems Science and Complexity*, 2017, **30**(3): 595–615.

[6] Xie S and Guo L, Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition, *SIAM Journal on Control and Optimization*, 2018, **56**(5): 3404–3431.

[7] Gan D and Liu Z, Strong consistency of the distributed stochastic gradient algorithm, *Proceedings of the* 58*th Conference on Decision and Control*, Nice, France, 2019, 5082–5087.

[8] Xie S, Zhang Y, and Guo L, Convergence of a distributed least squares, *IEEE Transactions on Automatic Control*, DOI: 10.1109/TAC.2020.3047989.

[9] Gan D, Xie S, and Liu Z, Stability of the distributed Kalman filter using general random coefficients, *Science China Information Sciences*, 2021, **64**: 172204.

[10] Kar S, Moura J, and Ramanan K, Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication, *IEEE Transactions on Information Theory*, 2012, **58**(6): 3575–3605.

[11] Zhu S, Soh Y, and Xie L, Distributed parameter estimation with quantized communication via running average, *IEEE Transactions on Signal Processing*, 2015, **63**(17): 4634–4646.

[12] Zhu S, Liu S, Soh Y, et al., Performance analysis of averaging based distributed estimation algorithm with additive quantization model, *Automatica*, 2017, **80**: 95–101.

[13] Zhu S, Chen C, Xu J, et al., Mitigating quantization effects on distributed sensor fusion: A least squares approach, *IEEE Transactions on Signal Processing*, 2018, **66**(13): 3459–3474.

[14] Wang Y and Zhang J F, Distributed parameter identification of quantized ARMAX systems, *Proceedings of the* 38*th Chinese Control Conference*, 2019, 1701–1706.

[15] Tan S, Guo J, Zhao Y, et al., Adaptive control with saturation-constrainted observations for drag-free satellites — A set-valued identification approach, *Science China Information Sciences*, 2021, **64**(10): 202202:1–202202:12.

[16] Brailsford A, Yussouff M, and Logothetis E, Theory of gas sensors, *Sensors and Actuators B: Chemical*, 1993, **13**(1): 135–138.

[17] Wang L Y, Zhang J F, and Yin G, System identification using binary sensors, *IEEE Transactions on Automatic Control*, 2003, **48**(11): 1892–1907.

[18] Wang L Y and Yin G, Asymptotically efficient parameter estimation using quantized output observations, *Automatica*, 2007, **43**(7): 1178–1191.

[19] Godoy B, Goodwin G, Agüero J, et al., On identification of FIR systems having quantized output data, *Automatica*, 2011, **47**(9): 1905–1915.

[20] Guo J and Zhao Y, Recursive projection algorithm on FIR system identification with binary-valued observations, *Automatica*, 2013, **49**: 3396–3401.

[21] Wang T, Hu M, and Zhao Y, Convergence properties of recursive projection algorithm for system identification with binary-valued observations, *Proceedings of the* 2018 *Chinese Automation Congress*, 2018, 2961–2966.

[22] Guo J and Zhao Y, Identification of the gain system with quantized observations and bounded persistent excitations, *Science China Information Sciences*, 2014, **57**: 1–15.

[23] Zhang H, Wang T, and Zhao Y, Asymptotically efficient recursive identification of fir systems with binary-valued observations, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, **51**(5): 2687–2700.

[24] Csáji B C and Weyer E, Recursive estimation of ARX systems using binary sensors with adjustable thresholds *IFAC Proceedings Volumes*, 2012, **45**(16): 1185–1190.

[25] Song Q, Recursive identification of systems with binary-valued outputs and with ARMA noises, *Automatica*, 2018, **93**: 106–113.

[26] Xiao J and Song Q, Recursive identification of quantized linear systems, *Journal of Systems Science and Complexity*, 2019, **32**(4): 985–996.

[27] Zhang L, Zhao Y, and Guo L, Identification and adaptation with binary-valued observations under non-persistent excitation condition, arXiv: 2107.03588, 2021.

[28] Liu Y, Dong X, Li Q, et al., Integrated strategy for fault-tolerant flight control system design with rate saturating actuators, 2019 *IEEE* 15*th International Conference on Control and Automation*, 2019, 136–141.

[29] Godsil C and Royle G, *Algebraic Graph Theory*, Spring-Verlag, Berlin, 2014.

[30] Calamai P H and Moré J J, Projected gradient methods for linearly constrained problems, *Mathematical Programming*, 1987, **39**: 93–116.

[31] Lai T L and Wei C Z, Least squares estimates in stochastic regression models with applications to identification and control dynamic systems, *Annals of Statistics*, 1982, **10**(1): 154–166.

[32] Chen H F and Guo L, *Identification and Stochastic Adaptive Control*, Birkhäsuser, Boston, 1991.

[33] Guo L, *Time-Varying Stochastic Systems Stability, Estimation and Control*, Second Edition, Science Press, Beijing, 2020.

[34] Jing L and Zhang J F, Tracking control and parameter identification with quantized ARMAX systems, *Science China Information Sciences*, 2019, **62**(9): 199203.

[35] Jing L and Zhang J F, LS-based parameter estimation of DARMA systems with uniformly quantized observations, *Journal of Systems Science and Complexity*, 2021, DOI: 10.1007/s11424-021-0314-y.